## ABSTRACT OF THE DISCLOSURE

Method and apparatus for content based HyperText Markup Language (HTML) coding. A computing device such as a server may include an HTML simplifier, encoder, and memory. Source HTML data may be accessed that presents HTML web pages. The HTML data is simplified. The simplification minimizes the size of the HTML data. Prior knowledge of the HTML data may be used during the simplification. The simplified HTML data is encoded using a Huffman code. The encoded/compressed HTML data is stored. HTML data is encoded/compressed based on an empirical study of common HTML pages, therefore, providing optimization and performance gains.